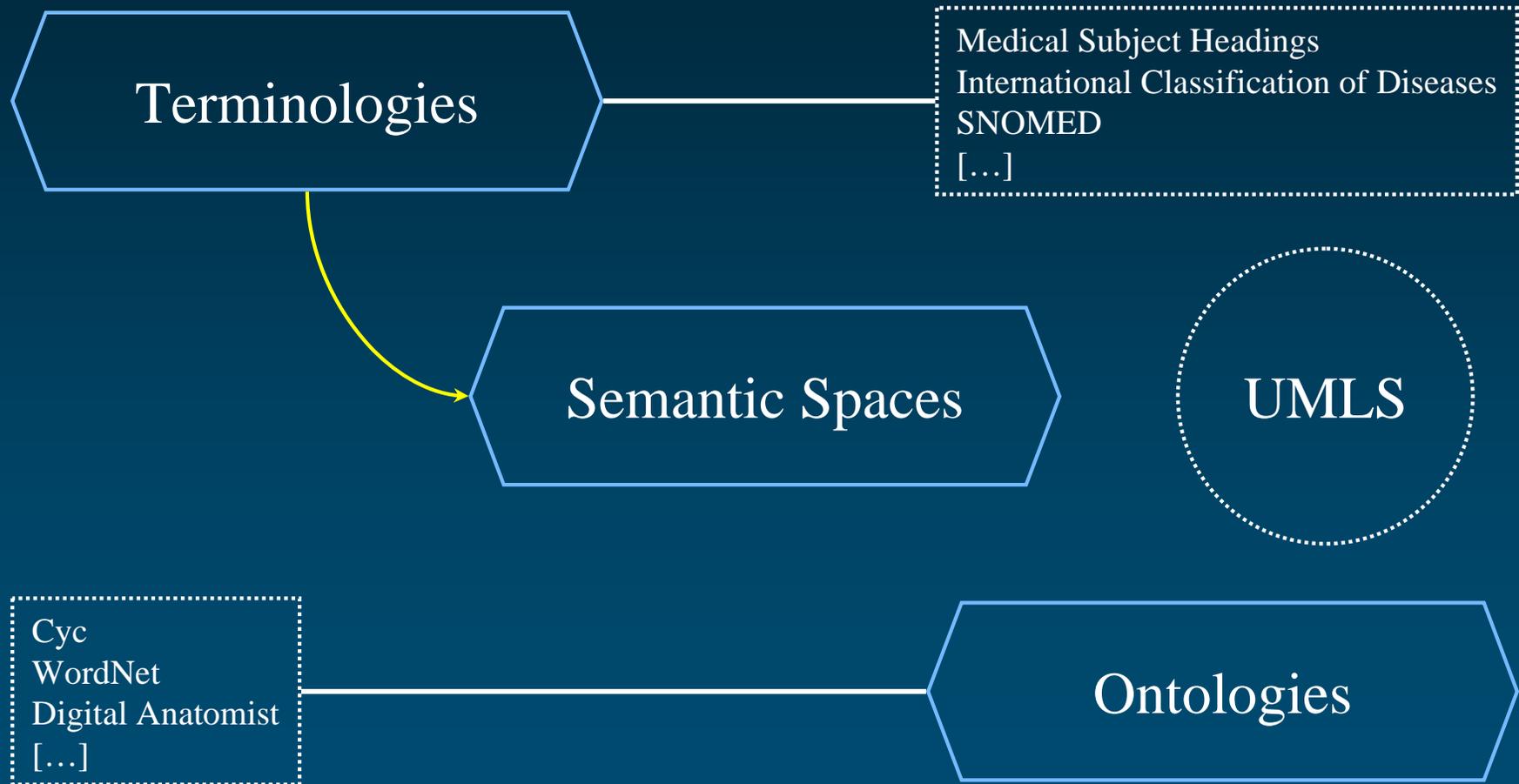


Final Report CgSB - July 11, 2001

*An ontological perspective
on the UMLS*

Anita Burgun

MOR (BOSC, May 17, 2001)



Introduction

- ◆ Semantics of the relationships between UMLS co-occurring concepts
- ◆ Semantic Grouping
- ◆ Lexical techniques for identifying hyponymic relations among medical terms
- ◆ Ontological features of the biomedical domain

UMLS Semantic space or ontology?

◆ Metathesaurus

- Concepts
- Relationships
- Huge
- Integrates vocabularies from other institutions
- Not censored



Semantic Space

◆ Semantic Network

- “a basic ontology for the biomedical domain”
- Semantic types
- Relationships
- Small
- Developed internally
- Fully controlled



Ontology

Semantic Network as domain ontology

- ◆ Re-used in specific medical areas:
 - MENELAS, MAOUSSC, genomics (Yu & al), blood transfusion (Achour & al)
- ◆ Re-used for specific tasks
 - Semantic tagging of medical documents (MEDTAG)
 - Natural Language Processing (Semantic Interpretation)
- ◆ Integrated into large-scale ontology libraries
 - ONIONS
- ◆ Addressed some ontological issues, e.g., polysemy

Semantic Network as domain ontology

- ◆ Systematic approaches for analyzing the UMLS
 - Structural: object-oriented model (Perl)
 - Semantic: Semantic Grouping
 - Ontological

Overview

- ◆ Illustrate our work from examples rather than report on all aspects
- ◆ 3 aspects:
 - The principles and the UMLS
 - Compatibility with a general ontology (WordNet)
 - Discussion inspired by the representation of the biomedical domain in several systems (“Blood”)

The principles and the UMLS

1- Formal properties

Formal properties

◆ Rigidity

- property that is essential to all the instances.
Person (+R). Physician (not R).

◆ Identity

- there is a property that is both necessary and sufficient for identifying an instance. Person (+I)

◆ Unity

- instances are intrinsic wholes. Person (+U).

◆ Dependence

- for all the instances x, necessarily some instance of Z must exist, which is not a part of x, nor a constituent of x (+D). Food (+D)

Formal properties Rules

◆ Rules

- (not U) cannot subsume (+U)
e.g., Substance cannot subsume Physical Object
- [...]

◆ Distinction between roles and sortal types

- Roles: (Not Rigid) (+Dependent)
- Sortal types : (+Rigid) (Not Dependent)

Formal properties Examples

- ◆ **Signs or Symptoms** are Roles
- ◆ **Eye Symptom** would belong to the SN
- ◆ Metathesaurus concepts that are assigned only to roles with no sortal Semantic Type represent a numerous set of entities (95% of the **Findings**, 86% of the **Signs or Symptoms** are not assigned to another Semantic Type).
- ◆ Which sortal type for : Heart murmur, innocent, Overactive child, or Early waking ?

The principles and the UMLS

2- The Economy Principle

The economy principle

◆ R1. Ad hoc precision

- The intent is to establish a set of semantic types, which will be useful for a variety of tasks without introducing undue complexity. The most specific semantic type in the semantic type hierarchy is assigned to the concept.

◆ R2. No hybrid types

- Instead of creating a lattice structure, with hybrid types inheriting from two supertypes, the SN has a single inheritance tree structure. As a consequence, a Metathesaurus concept inheriting from two STs is assigned to both types.

◆ R3. No category “other”

- Rather than proliferating the number of semantic types to encompass multiple additional subcategories, concepts that cannot be categorized by any sibling Semantic Type are simply assigned their common supertype.

The economy principle and the theory

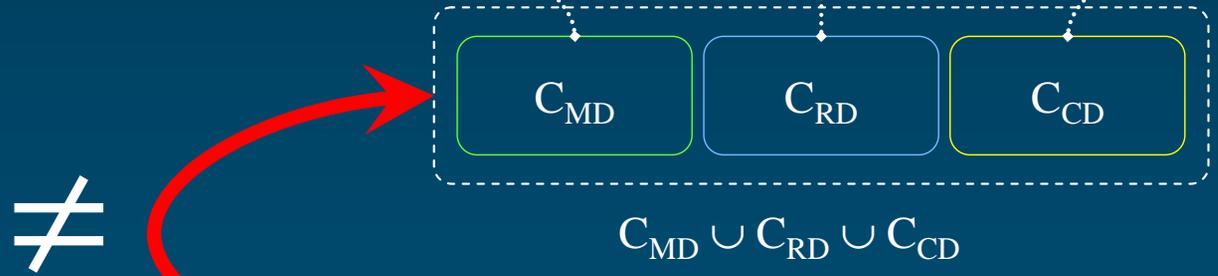
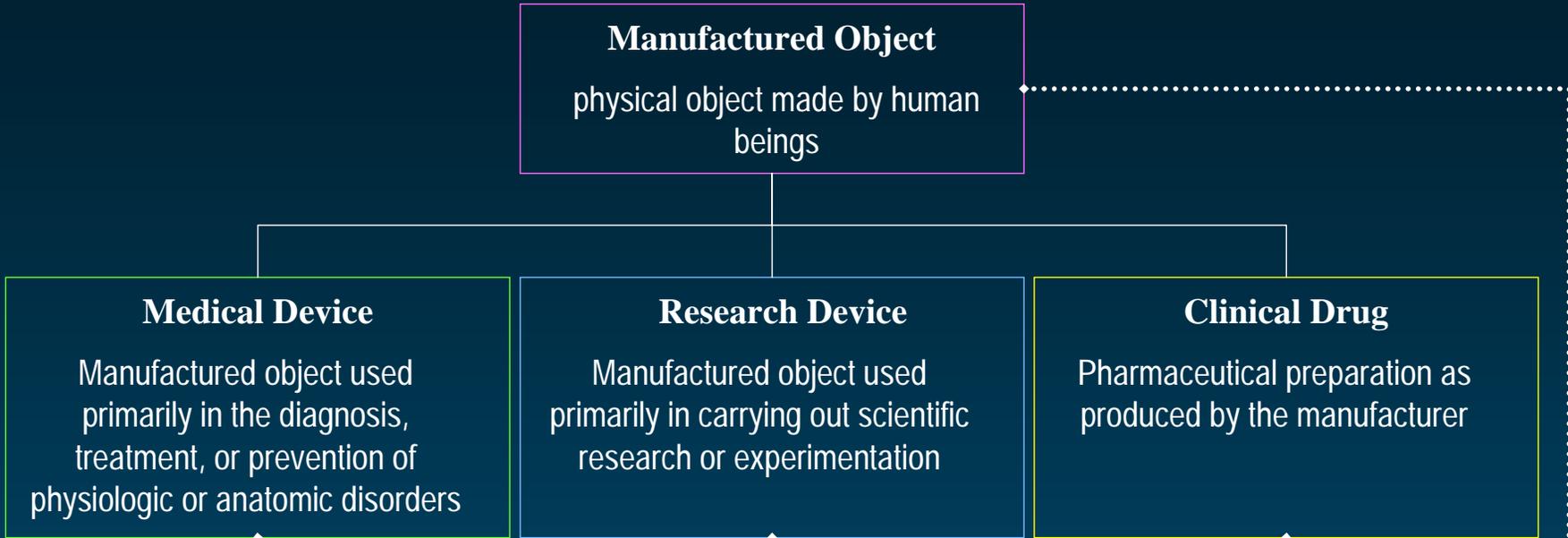
◆ Intensions and extensions

- Taxonomies (isa) are systems in which categories (intensions) are related to one another by means of subordination, or, in class parlance (extensions), systems in which classes are related to one another by means of class inclusion.

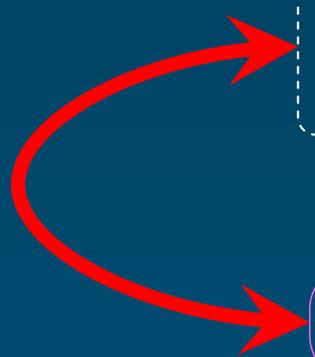
◆ Categories and classes

- When a category K has subcategories K_1, K_2, \dots, K_n , its extension, the class C_K is the union of the classes for each of its subcategories, i.e. $C_{K_1}, C_{K_2}, \dots, C_{K_n}$.

Categories

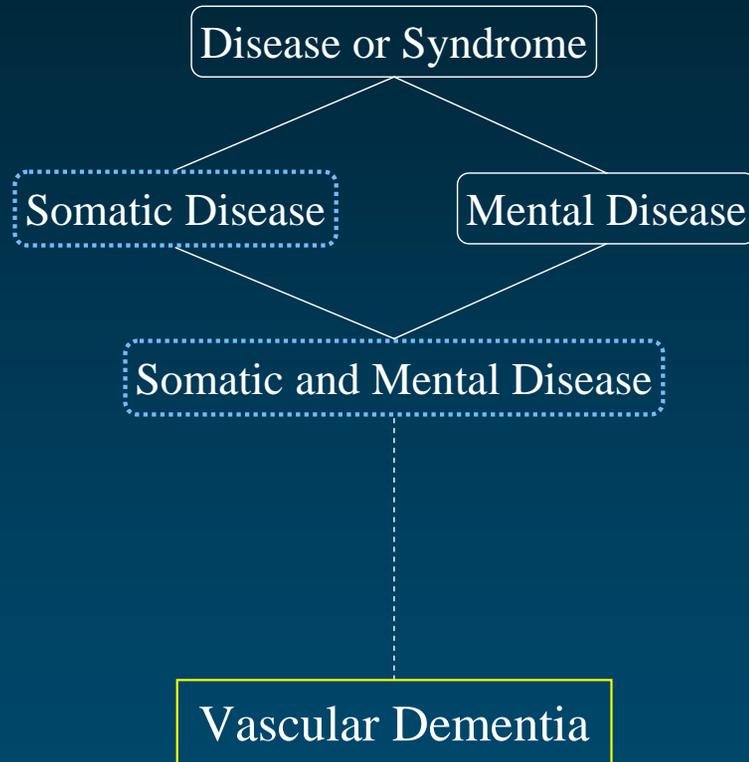


\neq



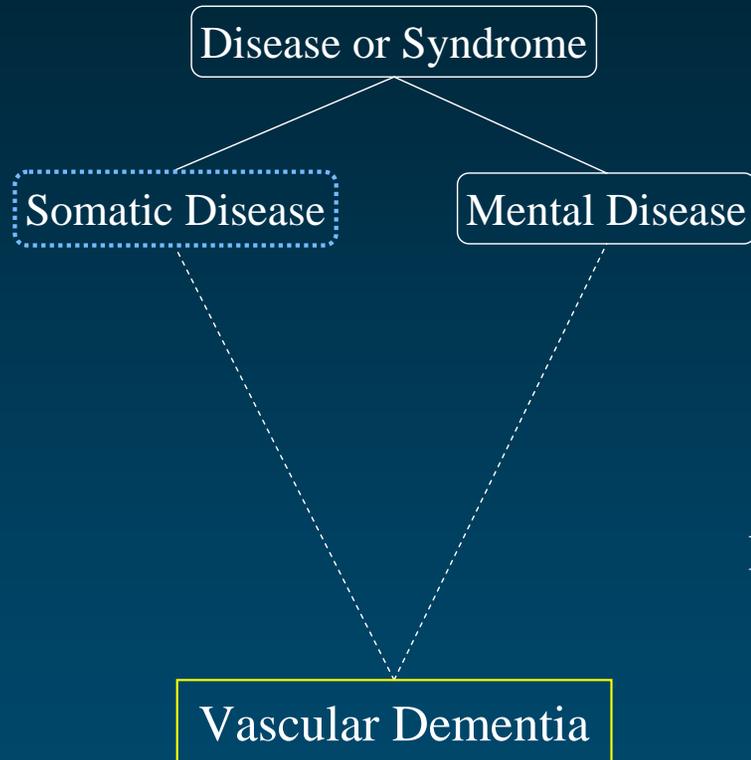
Classes

The economy principle and the theory



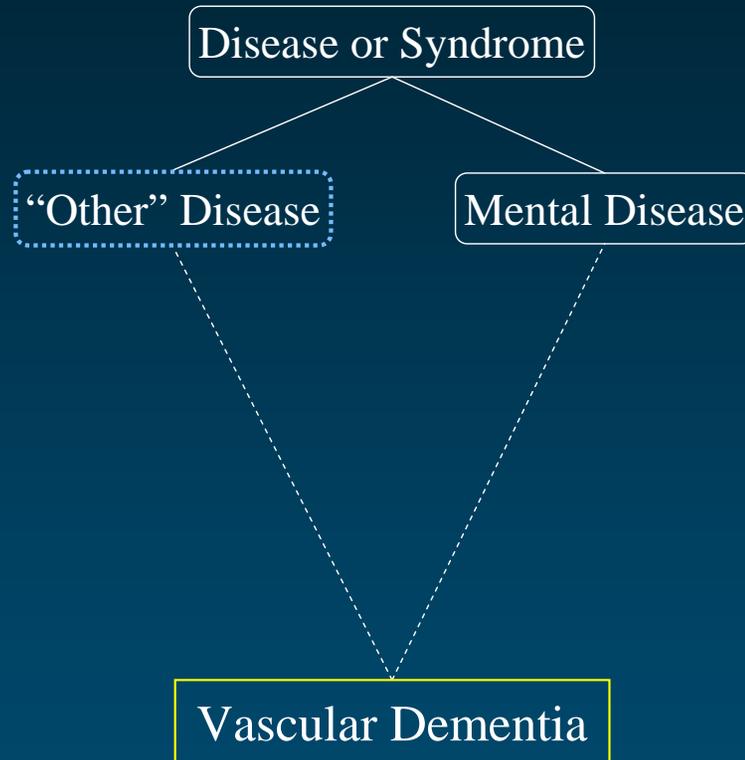
Rule R2: No hybrid type
Thus multiple categorization

The economy principle and the theory



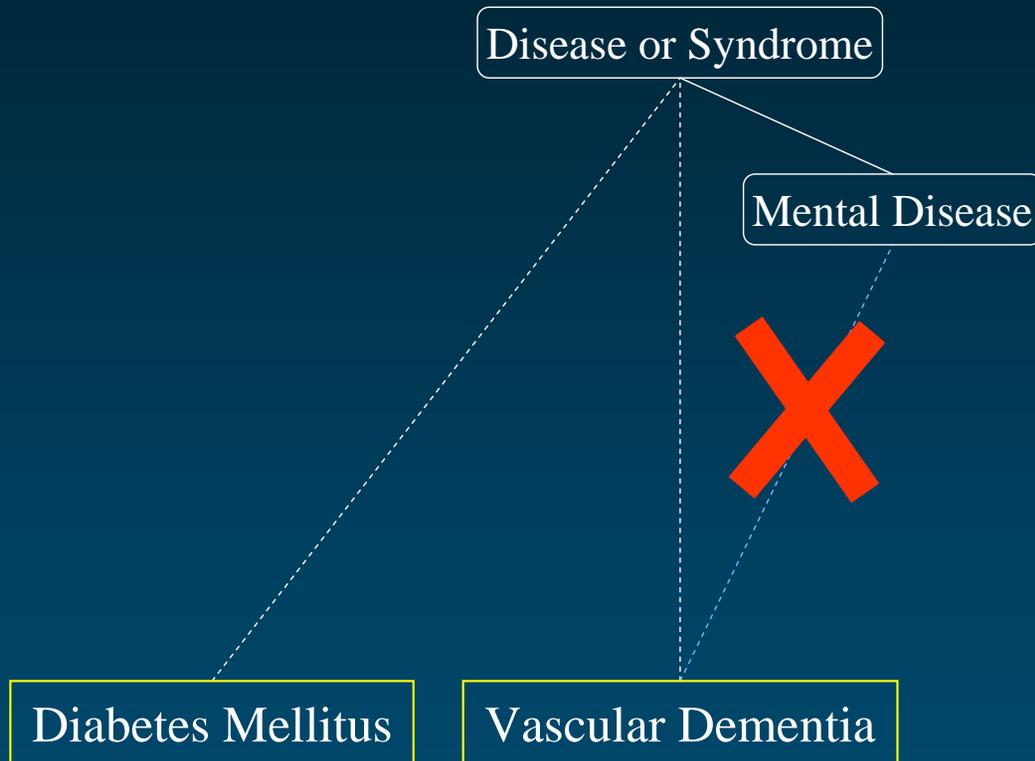
Rule R1: ad hoc precision in hierarchies
Thus no “Somatic disease” type

The economy principle and the theory



Rule R3: No “other” type
Assign to the common supertype

The economy principle and the theory

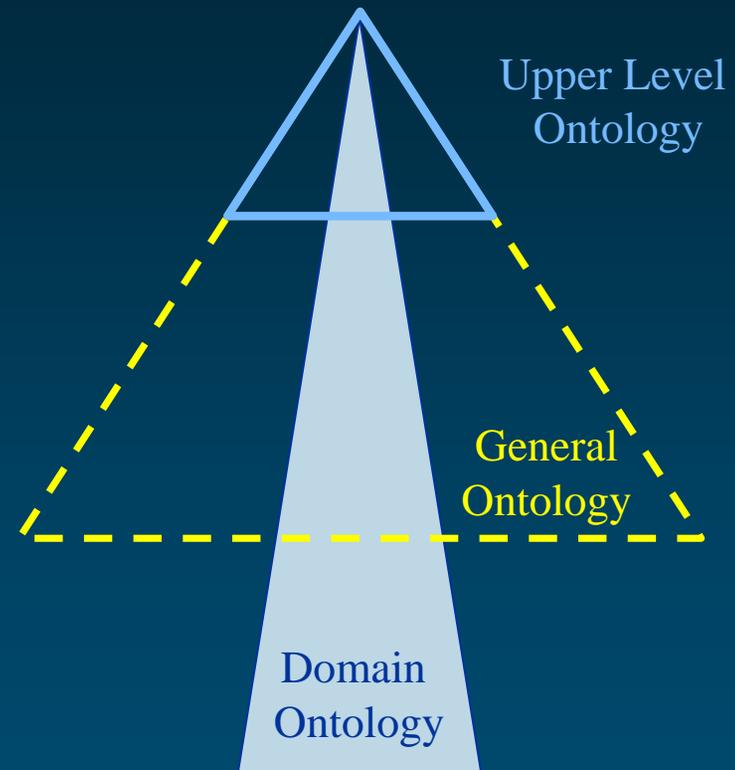


Compatibility with general ontologies

The example of WordNet

Compatibility among ontologies

- ◆ Compatibility in depth
 - Lower levels of ULO/
domain categories
(e.g., Disease)
- ◆ Compatibility in breadth
 - Categories that do not
specifically belong to D
(e.g., Manufactured Object)
- ◆ Universal Compatibility
 - Generic theories
(e.g., time, space)
 - Meta-level categories
(e.g., properties, roles)



UMLS and WordNet (general ontology)

◆ WordNet

- Electronic lexical database (Princeton)
- General world; 100,000 synsets (clusters of terms)

◆ 3 levels

- Terms: Does the term T from S1 also belong to S2?
- Concepts: How do terms for concept C in S1 overlap with terms for concept C' in S2?
- Semantic classes: How do concepts for class K in S1 overlap with concepts for class K' in S2?

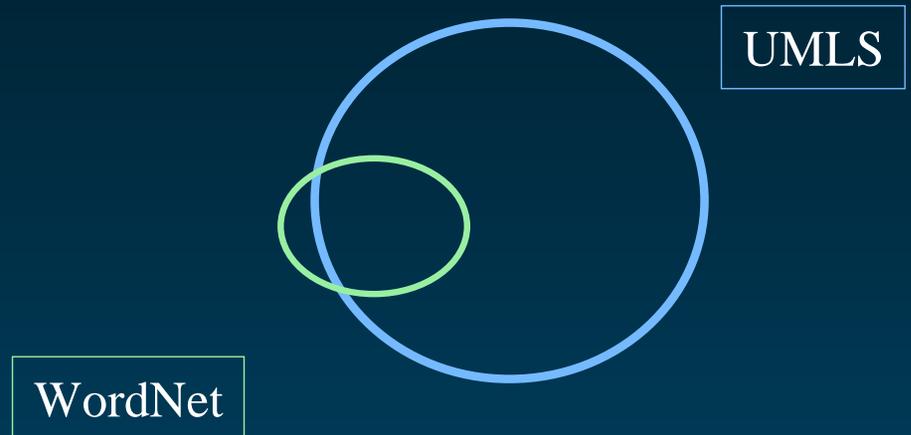
UMLS / WordNet Methods

- ◆ Compatible structures: Clusters of synonymous terms (concepts/ synsets)
- ◆ *isa* relations
 - UMLS: categorization
 - Health Disorder : Semantic Group Disorder
 - WordNet: hyponymy
 - Hyponyms of selected synsets (*Symptom, Ill Health, Disorder (sense 1), Mental retardation, Mental Illness, Defect (sense 1), Abnormalcy*)

UMLS / WordNet Results

◆ Health disorders

- 2% of UMLS concepts found in WordNet
- 83% of WordNet synsets found in the UMLS



	From WordNet	Found in UMLS
Synsets	1,379	83%
Terms	2,194	77%

Same class: 97%

	From UMLS	Found in WordNet
Concepts	143,991	2%

Same class: 48%

Specific terms

◆ UMLS

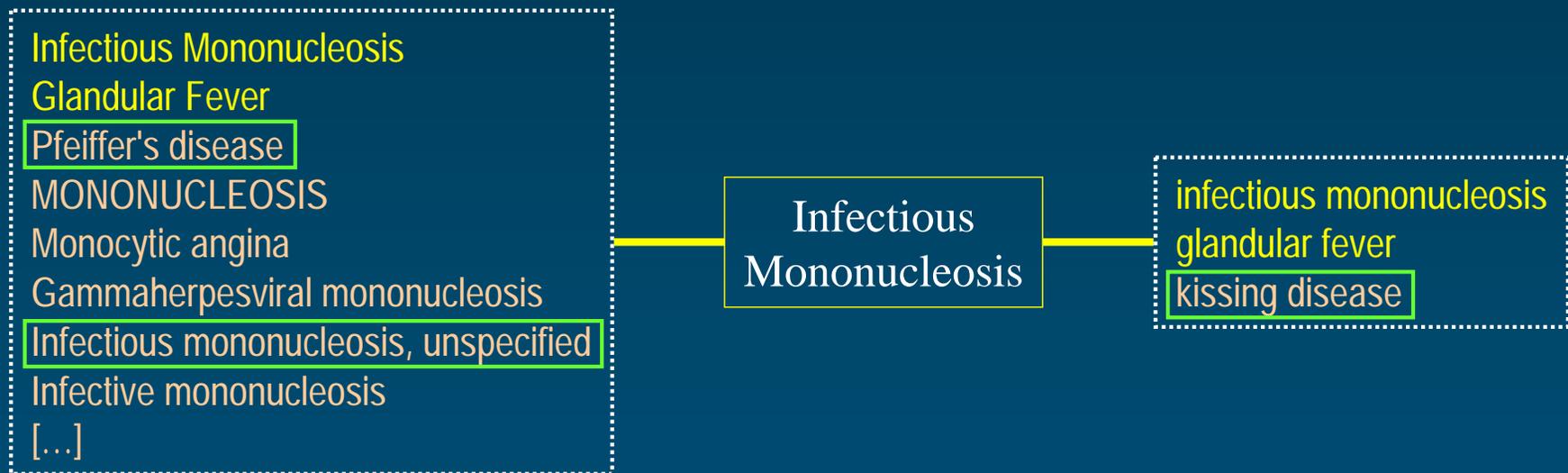
- Specialized terms
- Terminology-specific terms

◆ WordNet

- Lay synonyms

UMLS

WordNet



Specific concepts

◆ UMLS

- Health disorder
 - Many domain-specific concepts

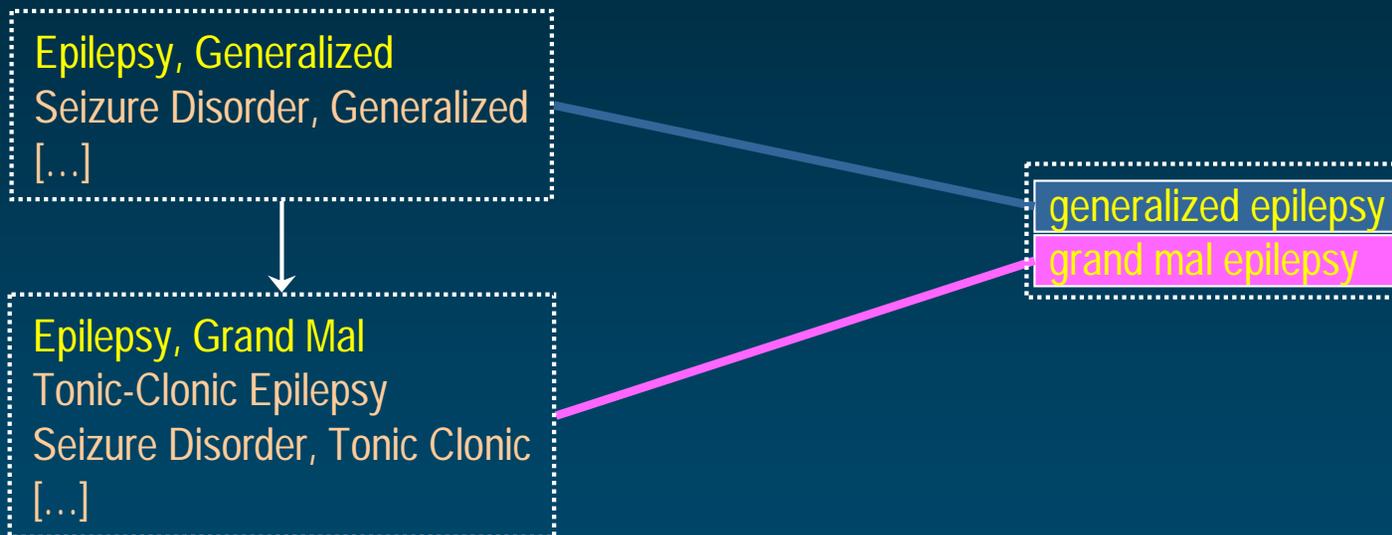
◆ WordNet

- Health disorder
 - Plant diseases
 - Astraphobia
 - Crick
 - Sword cut

Granularity, plesionymy

UMLS

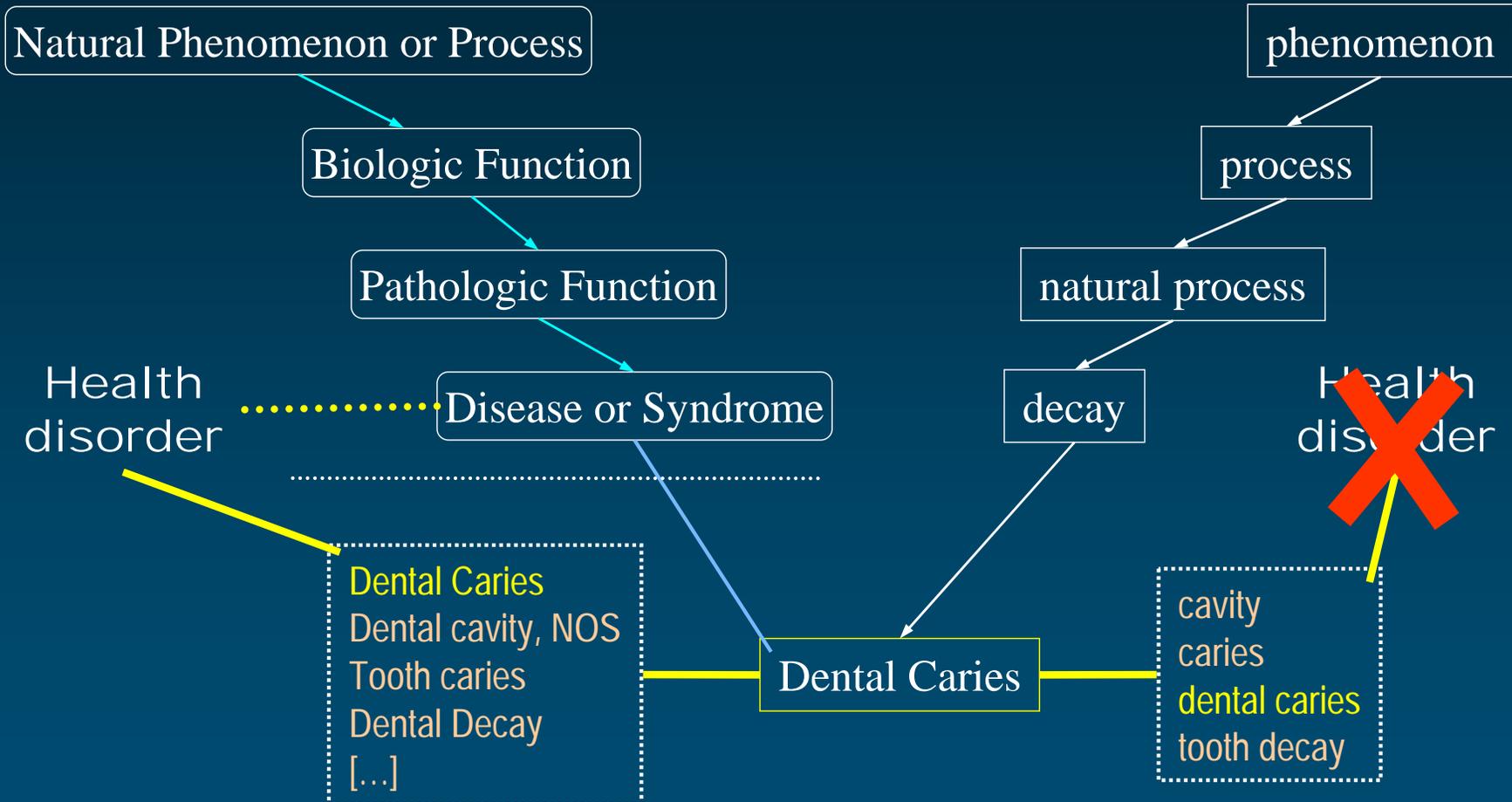
WordNet



Differing categorization

UMLS

WordNet



Representation of the biomedical domain in different systems

Blood

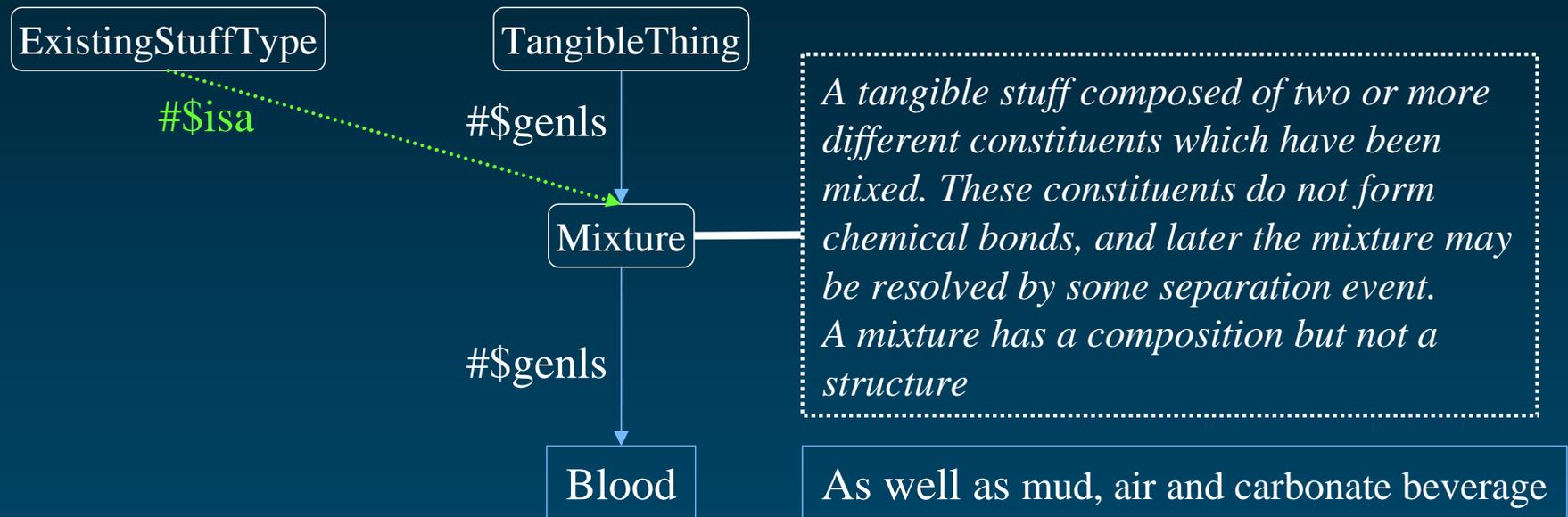
Objective

- ◆ Analyze core categories central to the biomedical domain
 - E.g., anatomy
- ◆ Study specific views, respective contributions of each system

Representation of Blood

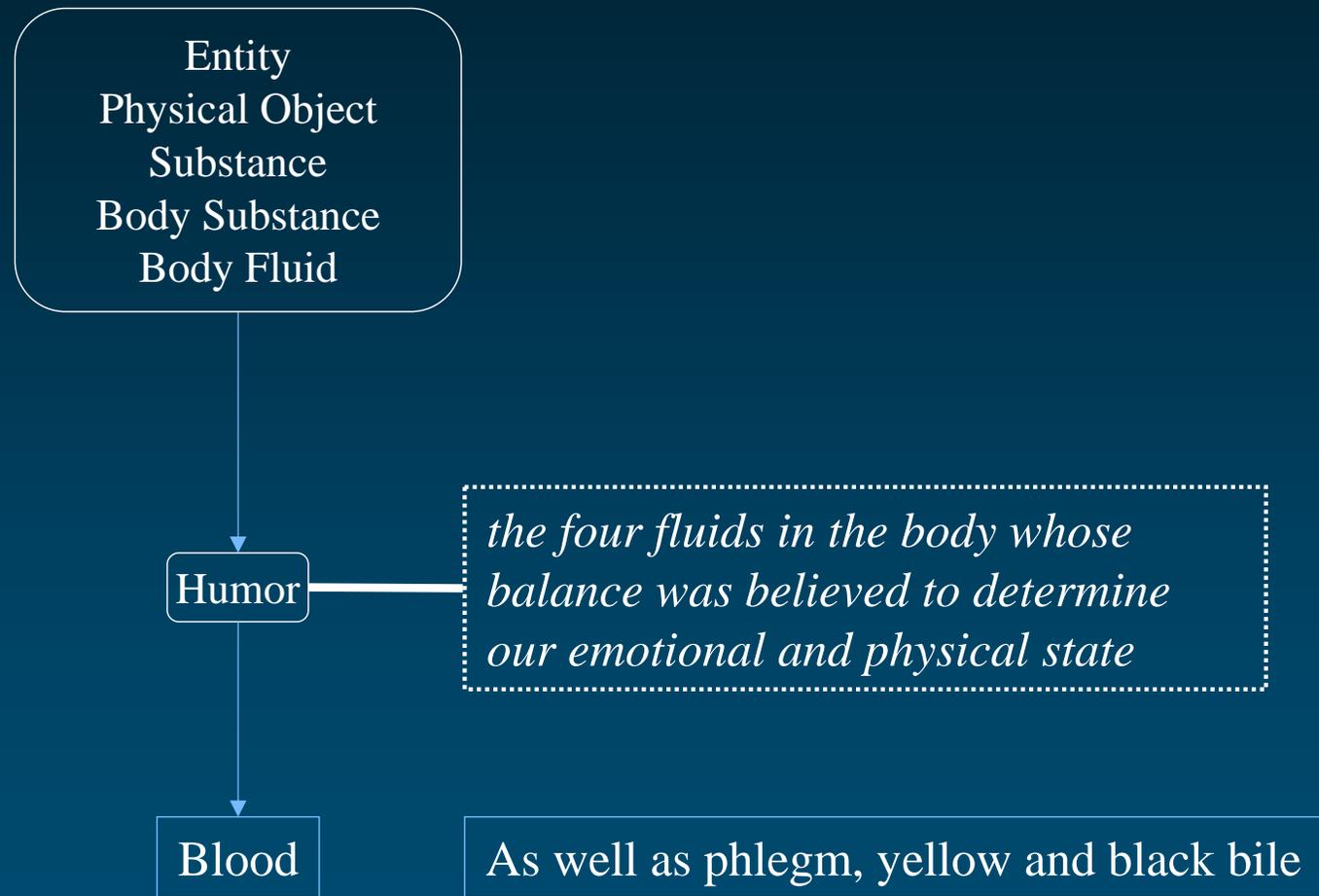
- ◆ In general ontologies
 - Cyc Knowledge Representation, common-sense
 - WordNet
- ◆ In domain ontologies
 - GALEN
 - UMLS
- ◆ In a specific ontology : Digital Anatomist
- ◆ In application ontologies : MENELAS

Cyc

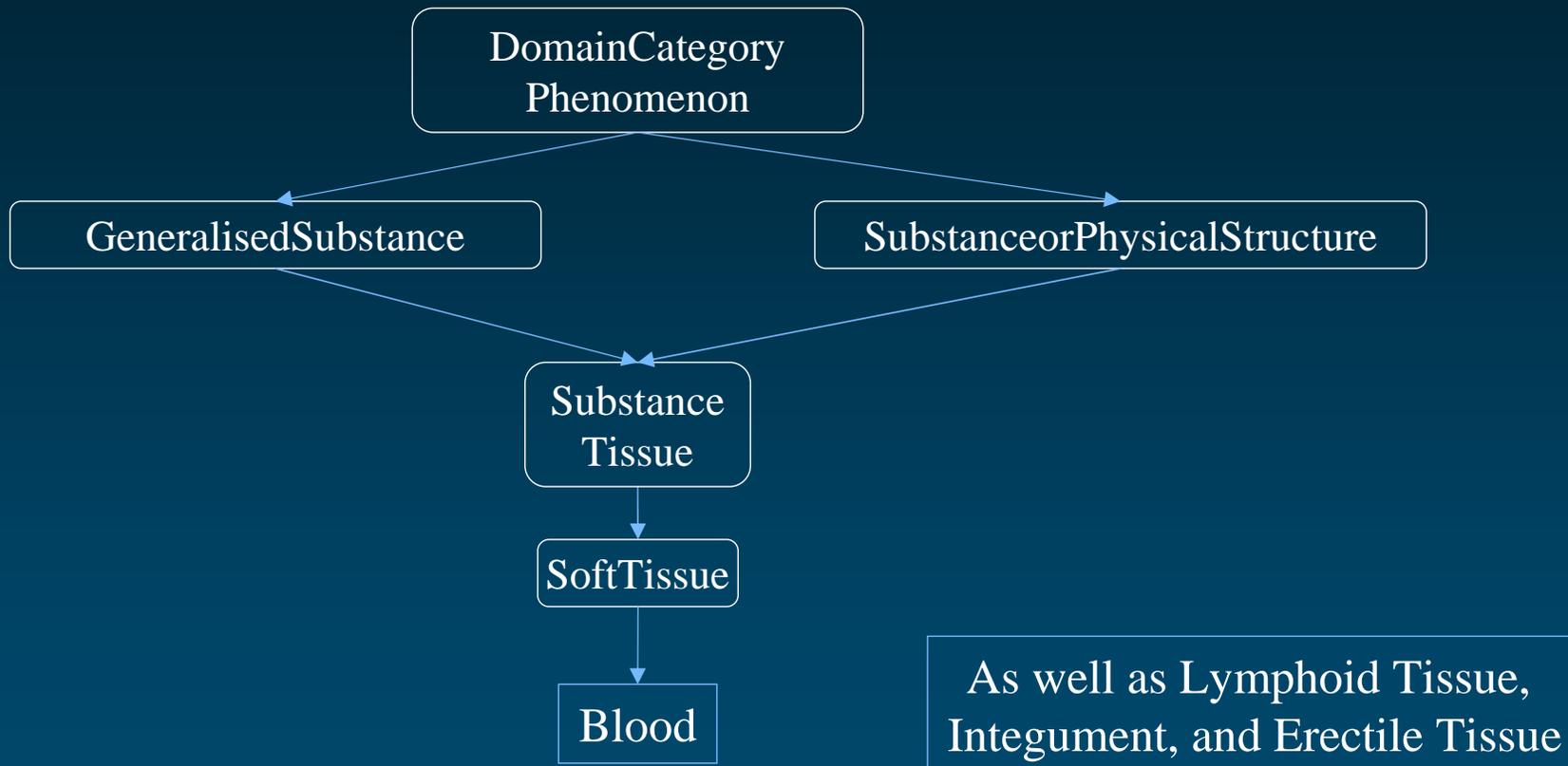


The function Separation-Event can apply to it.

WordNet

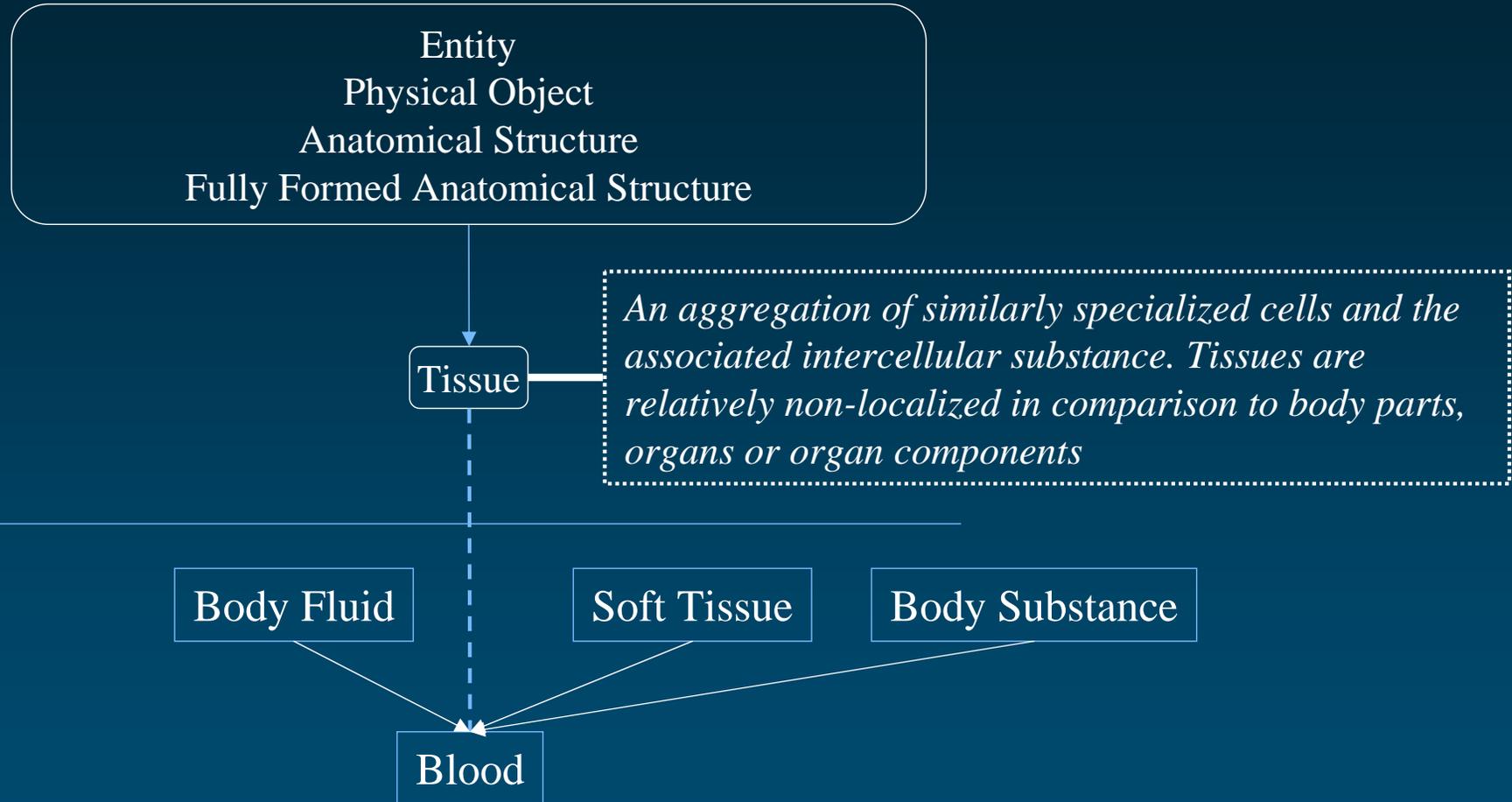


GALEN



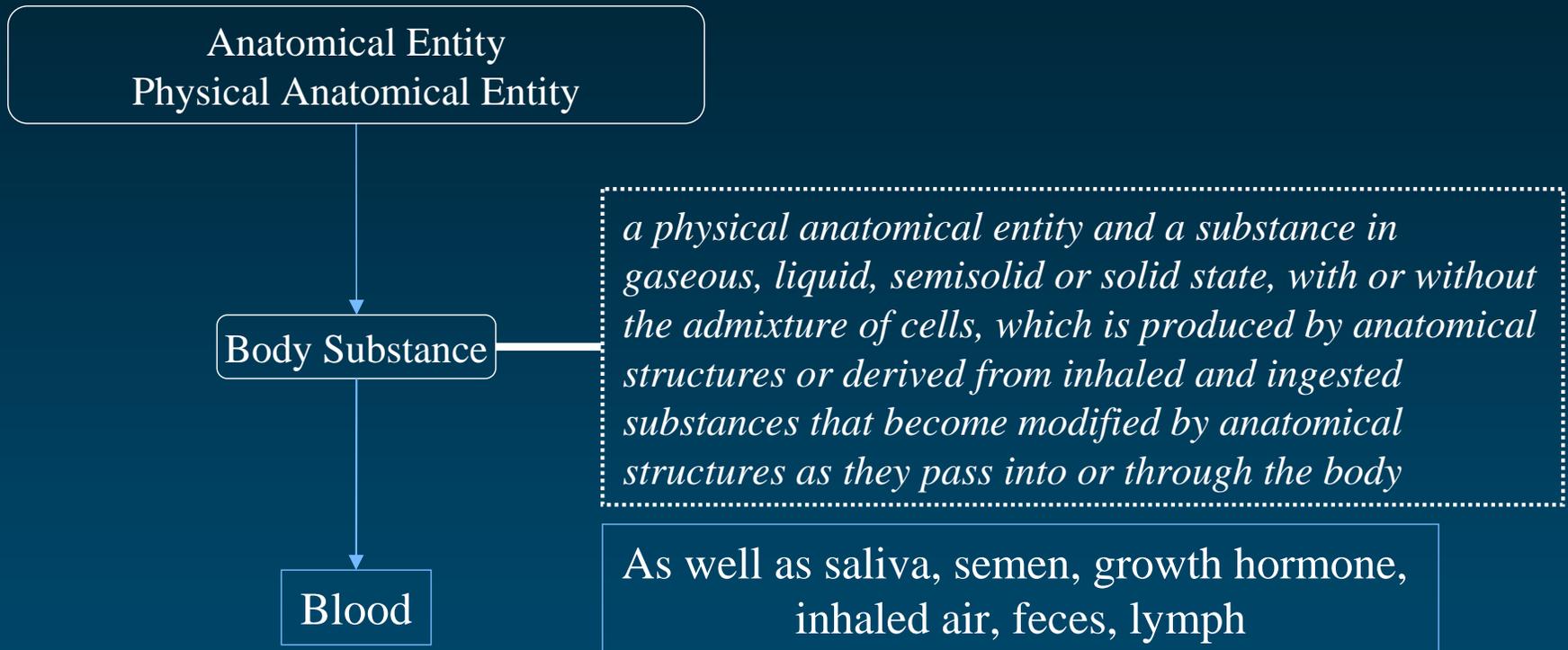
Blood has two states, `LiquidBlood` and `CoagulatedBlood`

UMLS



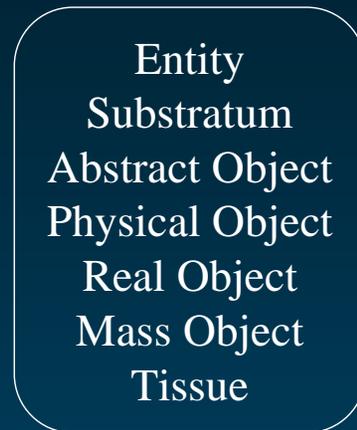
Tissue Produces Biologically Active Substance

Digital Anatomist



Tissue is an Organ Part.

MENELAS



Body Fluid

Blood

Model `body_fluid(_x)` is
`[body_fluid: _x]--(attr)-->[viscosity]`

As well as Lymph

Mass Objects are constituted of Countable Objects

From an example to discussion about...

- ◆ Knowledge and representations of knowledge
 - Within the biomedical domain (core concepts)
 - Definition of Tissue
 - Expert knowledge vs. general
 - Humors as microtheories
 - Upper level categories
 - Mixtures in Cyc, Mass objects (non countable) in MENELAS
 - Level of Knowledge to be represented in a DO
 - Coagulated Blood, Liquid Blood in GALEN

Future Plans

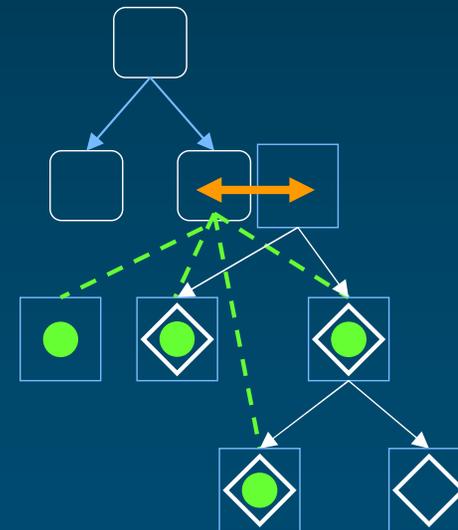
Future plans

◆ Several projects with CgSB

- Alignment Metathesaurus/ SN (the descendants of A/ the Metathesaurus concepts assigned to A)
- Comparing definitions in WordNet and the UMLS
- Findings?

◆ Formal aspects

- N. Guarino



Acknowledgments

- ◆ Alexa McCray
- ◆ Olivier Bodenreider
- ◆ Tom Rindflesch



Contact: Anita.Burgun@univ-rennes1.fr